



Measurement of the Top Quark Mass in the Dilepton Channel using a Matrix Element Method and Neuroevolution Selection with 2.0 fb^{-1}

The CDF Collaboration
URL <http://www-cdf.fnal.gov>
(Dated: December 5, 2007)

We report a measurement of the top quark mass using events collected by the CDF II Detector from $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$ at the Fermilab Tevatron. We calculate a likelihood function for the top mass in events that are consistent with $t\bar{t} \rightarrow \bar{b}\ell^-\bar{\nu}_\ell b\ell'^+\nu'_\ell$ decays. The likelihood is formed as the convolution of the leading-order matrix element and detector resolution functions. The events used in this analysis are selected using an evolutionary neural network that is directly optimized for precision in measurement of the top quark mass. The joint likelihood is the product of likelihoods for each of 344 events collected in 2.0 fb^{-1} of integrated luminosity, yielding a top quark mass, $M_t = 171.2 \pm 2.7(\text{stat.}) \pm 2.9(\text{syst.}) \text{ GeV}/c^2$.

I. INTRODUCTION

Precision measurements of the top quark mass, M_t , place constraints on the masses of particles to which the top quark contributes radiative corrections, including the unobserved Higgs boson and particles in extensions to the standard model. At the Tevatron, top quarks are primarily produced in pairs. The dilepton channel, consisting of the decays $t\bar{t} \rightarrow \bar{b}\ell^-\bar{\nu}_\ell b\ell'^+\nu'_\ell$, has a small branching fraction but allows measurements which are less reliant on the calibration of the jet energy scale, the dominant systematic uncertainty, than measurements in channels with hadronic W decays. A discrepancy from other channels could indicate contributions from new processes.

The reconstruction of the top mass from dilepton events poses a particular challenge as the two neutrinos from W decays are undetected. Previous measurements in this channel [1, 2] using Run I data have calculated the mass by making several kinematic assumptions and integrating over the remaining unconstrained quantity. To extract maximum information from the small sample of dilepton events, we adapt a technique pioneered for the analysis of $t\bar{t} \rightarrow b\ell\nu_\ell\bar{b}q\bar{q}'$ decays [3–5]. This technique uses the leading-order production cross-section and a parameterized description of the jet energy resolution. Making minimal kinematic assumptions and integrating over six unconstrained quantities, we obtain per-event likelihoods in top mass which can be directly multiplied to obtain the joint likelihood from which M_t is determined.

This method was first applied to dilepton events at CDF [6–8] using a data set of 340 pb^{-1} . An updated method

was applied to a data set of 1.8 fb^{-1} [9], yielding the single most precise measurement of M_t in the dilepton channel. These previous analyses utilized an event selection developed and optimized for measurement of the $t\bar{t}$ production cross section in the dilepton channel [10].

We present here a measurement performed using events selected with a neural network (NN) that is specifically optimized via an evolutionary approach for maximizing the precision in top quark mass measurements.

II. EVENT SELECTION

This section describes the collected dataset, the trigger, the topological pre-selection which forms the basis of the final optimized selection, and the optimization of the selection with neuroevolution.

A. Dataset and Trigger

This analysis is based on an integrated luminosity of 2.0 fb^{-1} collected with the CDF II detector between March 2002 and May 2007. The CDF II detector is a general purpose detector described elsewhere [12]. The data are collected with an inclusive lepton trigger that requires events to have a lepton with $E_T > 18 \text{ GeV}$ (for an electron) or $p_T > 18 \text{ GeV}/c$ (for a muon).

B. Pre-selection

We define a pre-selection which satisfies the basic topology of top dilepton decay and provides a starting point for the optimization of selection used in the mass measurement. The pre-selection requires events with i) two opposite sign leptons (electrons or muons) with $p_T > 20 \text{ GeV}/c$, ii) two or more jets with $E_T > 15 \text{ GeV}$ within the region $|\eta| < 2.5$, iii) missing transverse energy, $\cancel{E}_T > 20 \text{ GeV}$, and iv) the invariant mass of the two leptons, $M_{ll} > 10 \text{ GeV}/c^2$. We validate our pre-selection by comparing kinematic distributions for data events passing the pre-selection to those from simulated events processes predicted to be in the pre-selected sample as shown in Figure 1.

C. Optimization with Neuroevolution

Finding the selection which produces the smallest expected statistical uncertainty for the top mass is a non-trivial problem. The analysis technique itself has powerful mechanisms for suppressing certain backgrounds, which suggests the measurement would be improved by loosening selection cuts to allow more signal as well as more well-suppressed background. On the other hand, the likelihood calculations done in the analysis require several approximations which require a correction; events which violate these assumptions may have a negative effect on the sensitivity.

Neural networks can describe an arbitrary function, so that the problem of finding the optimal selection function can be recast as a search through the space of neural network weights and topologies for a network which gives the optimal selection. In high-energy physics, neural networks are traditionally trained to separate signal and background events in a way that minimizes the mis-classification of signal events as background and vice-versa. In this case, we cannot use the traditional training technique, because we do not seek to classify individual events, but instead to optimize the total analysis sensitivity, a property of an ensemble of events.

Instead, we use neuroevolution, an evolutionary approach to directly search for the optimal network. Beginning with a population of random networks, analysis sensitivity is evaluated for each network. As seen in Figure 2, poor performers are culled and strong performers are bred together and mutated in successive generations until performance reaches a plateau. Because we have optimized directly on the analysis sensitivity, rather than some intermediate or approximate figure of merit, the final network is precisely the one which gives the most sensitive measurement. This approach has been shown to significantly outperform supervised methods in top quark event selection [13]. In particular, we use NeuroEvolution of Augmenting Topologies (NEAT) [14], a neuroevolutionary method capable of evolving a network's topology in addition to its weights.

The neural network input variables are:

- Missing transverse energy
- Total vector energy
- Number of observed b -tags

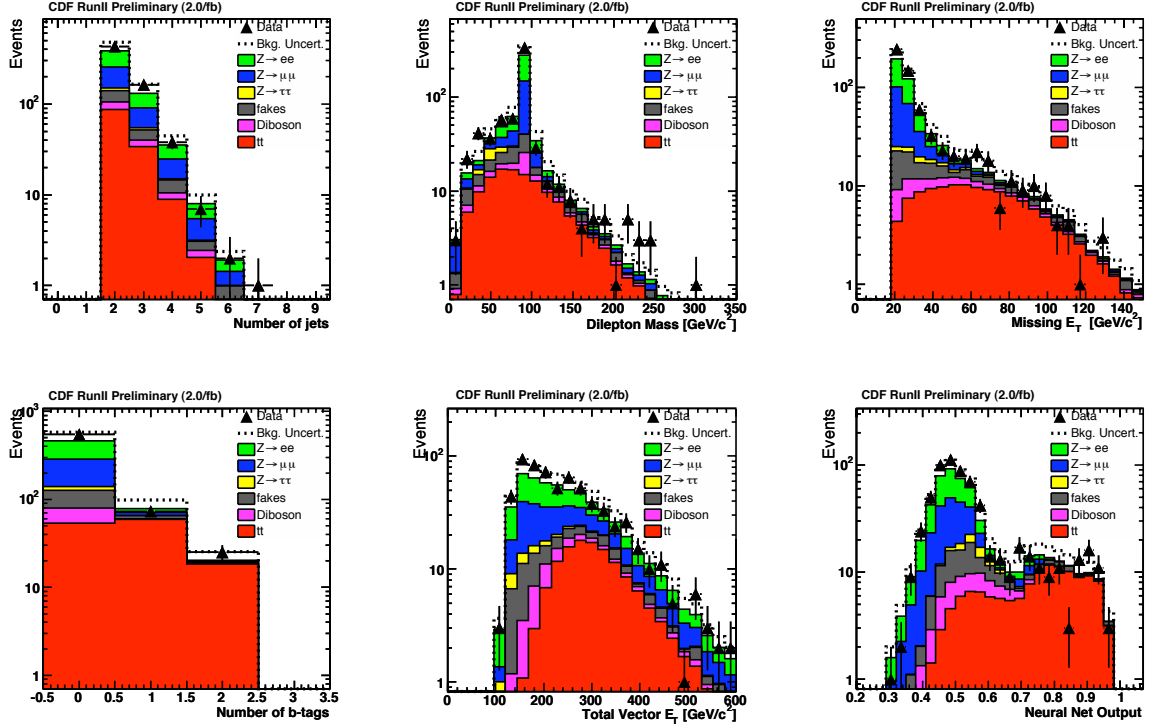


FIG. 1: Distribution of kinematic variables after preselection: number of jets, invariant mass of leptons, missing transverse energy, number of secondary vertex tags, total vector E_T in the event, and the output of the NN used to optimize our selection.

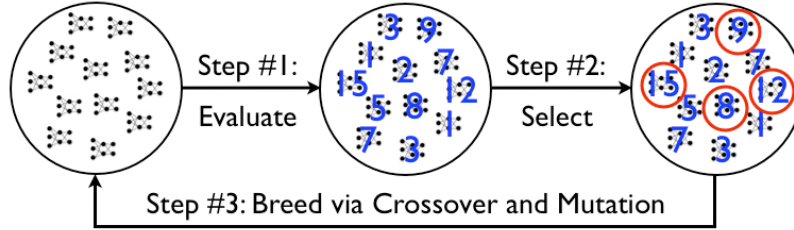


FIG. 2: Neuro-evolution: evolutionary search for the optimal neural network event selection. Left, a population of neural networks. Center, each network is evaluated in the context of the full analysis. Right, the networks with the best performance are used to found the next population, and the poor performers are discarded before the cycle repeats.

- Dilepton invariant mass
- Minimum azimuthal angle between a jet and the missing transverse energy
- Minimum azimuthal angle between a lepton and the missing transverse energy

We perform 20 generations of neuroevolution with this algorithm before choosing a neural network to select events for our analysis. The predicted number of events for each expected source for a 2.0 fb^{-1} is shown in Table I for both events with and without secondary vertex tags. Comparison of kinematic variables for data events passing the final NN selection and for simulated events processes predicted to be in the sample is shown in Figure 3. Using the NN optimized selection improves the *a priori* statistical uncertainty over the selection used in prior iterations of this analysis by 20%.

N_{tags}	0	≥ 1
$Z \rightarrow ee$	56.3 ± 13.7	2.3 ± 1.1
$Z \rightarrow \mu\mu$	48.6 ± 12.0	1.7 ± 0.8
$Z \rightarrow \tau\tau$	11.6 ± 4.0	0.1 ± 0.1
$Z \rightarrow ee + bb$	1.8 ± 0.6	4.3 ± 2.2
$Z \rightarrow \mu\mu + bb$	1.5 ± 0.5	3.4 ± 1.8
$Z \rightarrow \tau\tau + bb$	0.2 ± 0.1	0.2 ± 0.1
$Z \rightarrow ee + cc$	3.0 ± 0.9	1.2 ± 0.7
$Z \rightarrow \mu\mu + cc$	2.4 ± 0.7	1.0 ± 0.5
$Z \rightarrow \tau\tau + cc$	0.4 ± 0.2	0.0
WW	11.0 ± 5.5	0.4 ± 0.4
WZ	3.3 ± 1.7	0.2 ± 0.2
ZZ	2.3 ± 1.1	0.1 ± 0.1
$W\gamma$	0.7 ± 0.2	0.0
<i>fakes</i>	29.0 ± 8.7	4.5 ± 1.1
<i>tt</i>	43.8 ± 4.4	78.0 ± 6.2
Total	215.8 ± 21.9	97.5 ± 7.2
Data (2.0 fb^{-1})	246	98

TABLE I: Summary of events in as a function of b -tags for NN selection.

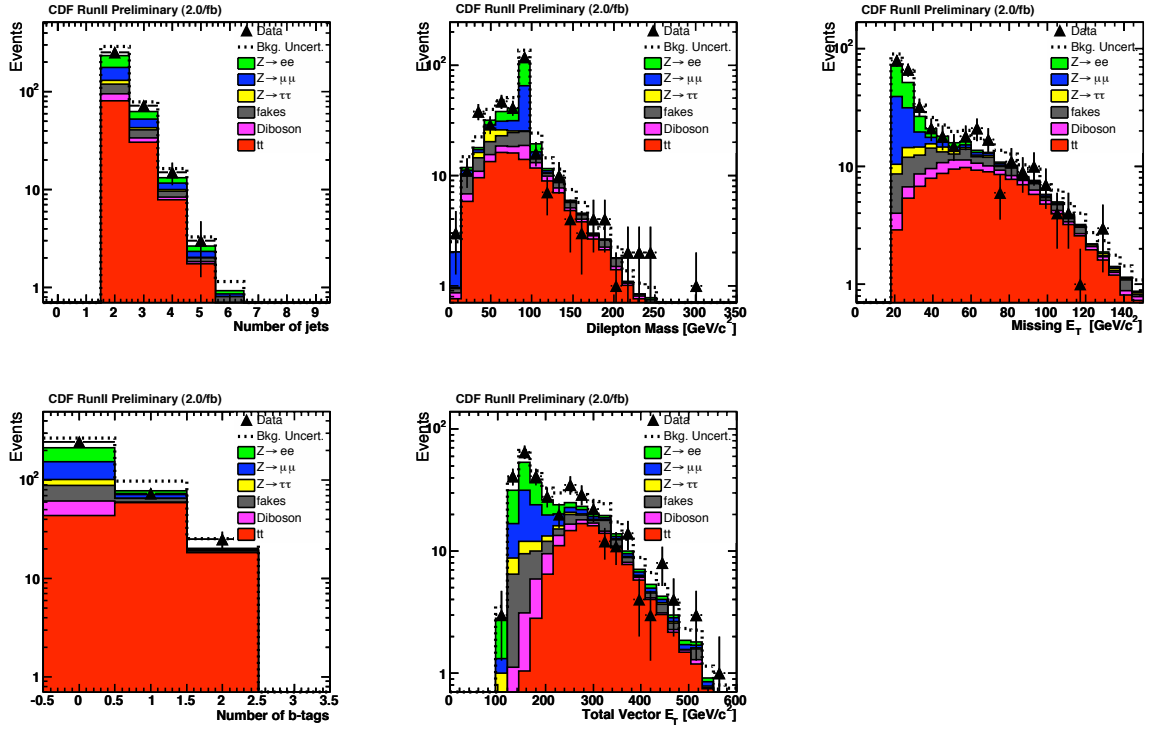


FIG. 3: Distribution of kinematic variables after preselection: number of jets, invariant mass of leptons, missing transverse energy, number of secondary vertex tags, and the total vector E_T in the event.

III. METHOD

The information contained in an event regarding the top mass can be expressed as the conditional probability $P(\mathbf{x}|M_t)$, where M_t is the top pole mass and \mathbf{x} is a vector of measured event quantities. We calculate the posterior probability using the theoretical description of the $t\bar{t}$ production process expressed with respect to the measured event

quantities:

$$P(\mathbf{x}|M_t) = \frac{1}{\sigma(M_t)} \frac{d\sigma(M_t)}{d\mathbf{x}}$$

where $\frac{d\sigma}{d\mathbf{x}}$ is the per-event differential cross-section.

To evaluate the probability, we integrate over quantities which are unknown because they are unmeasured by the detector, such as neutrino energies. Quark energies are not directly measured, but are estimated from the observed energies of the corresponding jets. We parametrize this uncertainty using a transfer function between quark and jet energies, $W(p, j)$, giving us the probability of measuring jet energy j given parton energy p . We form the transfer function by fitting a double Gaussian to a predicted distribution of parton-jet energy difference from simulated events. The total expression for the probability of a given pole mass for a specific event can be written as

$$P(\mathbf{x}|M_t) = \frac{1}{N} \int d\Phi_8 |\mathcal{M}_{t\bar{t}}(p; M_t)|^2 \prod_{jets} W(p, j) f_{PDF}(q_1) f_{PDF}(q_2) \quad (1)$$

where the integral is over the entire six-particle phase space, q is the vector of incoming parton-level quantities, p is the vector of resulting parton-level quantities: lepton and quark momenta, and $|\mathcal{M}_{t\bar{t}}(p; M_t)|$ is the $t\bar{t}$ production matrix element as defined in [15, 16]. The constant term in front of the integral ensures that the normalization condition for the probability:

$$\int d\mathbf{x} P(\mathbf{x}|M_t) = 1$$

is satisfied.

A. Background

The probability $P(\mathbf{x}|M_t)$ is sufficient to extract the top quark mass in an unpolluted sample. However, the event selection we utilize maximizes signal acceptance at the expense of a significant number of background events. To reduce the effect of these events on the measurement, we must calculate the probabilities, $P_{bg_i}(\mathbf{x})$ that they were produced by a given background process; we form the generalized per-event probability as

$$P^{n-tag}(\mathbf{x}|M_t) = P_s(\mathbf{x}|M_t)p_s^{n-tag} + P_{bg_1}(\mathbf{x})p_{bg_1}^{n-tag} + P_{bg_2}(\mathbf{x})p_{bg_2}^{n-tag} + \dots \quad (2)$$

simply a sum of the probabilities for each process, weighted by their respective priors. Here, $P_s(\mathbf{x}|M_t)$ is as described in equation 1 and the $P_{bg_i}(\mathbf{x})$ are formed by calculating a differential cross-section for each event in a manner similar to $t\bar{t}$. The background processes for which we evaluate probabilities for in this manner are: Drell-Yan with associated jets, W pair production with associated jets and $W+3$ jets production where one jet is incorrectly identified as a lepton.

The weights for each term in Equation 2 depend on whether the event has a secondary vertex tag and are determined from the expected fractions of signal and background events in each category, listed in Table I.

The final ensemble probability density is expressed as:

$$P(\mathbf{x}|M_t) = \left[\prod_{i_0} P_{0-t}(\mathbf{x}_{i_0}|M_t) \right] \times \left[\prod_{i_1} P_{1-t}(\mathbf{x}_{i_1}|M_t) \right], \quad (3)$$

where the products are over all untagged events, i_0 , and all tagged events, i_1 .

IV. CALIBRATION

To test the performance of the method, we construct Monte Carlo experiments using Monte Carlo for generated top masses from 155 GeV/ c^2 to 195 GeV/ c^2 . The number of signal and background events in each Monte Carlo experiment are Poisson fluctuated values around the *a priori* estimates given in Table I; the estimate for $t\bar{t}$ at varying masses is evolved to account for the variation of cross-section and acceptance. The response of the method for Monte Carlo experiments with both signal and background is shown in Figure 4. A correction, as derived from this response, is applied to the measured value in data.

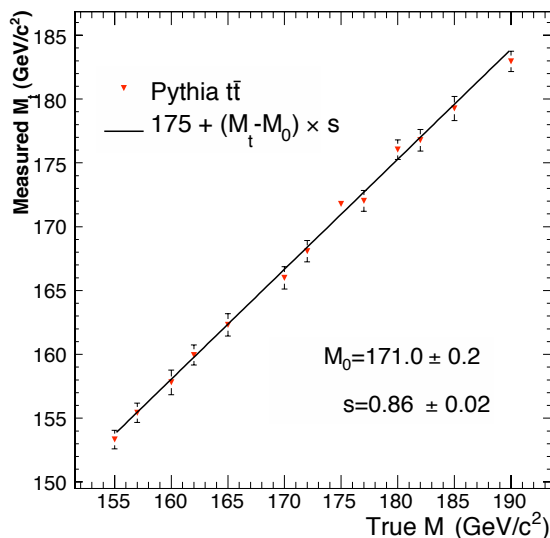


FIG. 4: Response for Monte Carlo experiments of signal and background events.

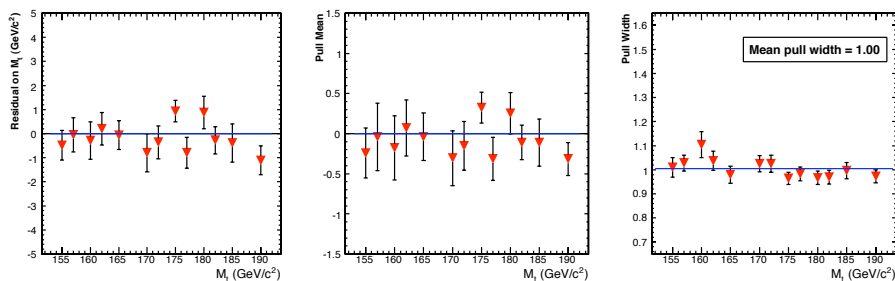


FIG. 5: Residual, pull mean, and pull width for varying top mass Monte Carlo samples.

In the interest of computational tractability, several assumptions are made in the evaluation of the integrals in Equation 1, such as the leading two jets in the event coming from b -quarks from top decay, and lepton momenta and jet angles being measured perfectly. These assumptions are violated in small and understood ways in realistic events. Due to these effects, the method underestimates the statistical error. The pull distribution in Monte Carlo experiments is shown in Figure 5. To account for this underestimation, we scale the statistical error by a factor, $S = 1.16$, derived from the results of our Monte Carlo experiments.

V. SYSTEMATIC UNCERTAINTIES

There are several sources of systematic error in our measurement which are summarized in Table II.

The single largest source of systematic error comes from the uncertainty in the jet energy scale, which we estimate by varying the jet energy corrections by $\pm 1\sigma$ and is $2.5 \text{ GeV}/c^2$. A small uncertainty is assigned to possible imperfections in the modeling of lepton energy in Monte Carlo events; this uncertainty is measured to be $0.1 \text{ GeV}/c^2$. The uncertainty in the Monte Carlo generator used to perform Monte Carlo experiments, estimated by measuring the difference in extracted the top mass from PYTHIA events and HERWIG events, amounts to $0.7 \text{ GeV}/c^2$. The uncertainty in the response correction shown in Figure 4 is estimated by varying that response by $\pm 1\sigma$ and is $0.4 \text{ GeV}/c^2$. The uncertainty due to initial-state (ISR) and final-state (FSR) radiation is estimated by varying the amount of ISR and FSR in simulated events and is measured to be $0.3 \text{ GeV}/c^2$ for both cases.

The uncertainty in background composition is estimated by varying the background estimates from Table I within their errors and amounts to $0.3 \text{ GeV}/c^2$. In addition, a large uncertainty comes from the limited number of Monte Carlo background events available for Monte Carlo experiments. To measure this uncertainty, we split each background

Source	Size (GeV/ c^2)
Jet Energy Scale	2.5
Lepton Energy Scale	0.1
Generator	0.7
Method	0.4
Sample composition uncertainty	0.3
Background statistics	0.5
Background modeling	0.2
FSR modeling	0.3
ISR modeling	0.3
PDFs	0.6
Total	2.9

TABLE II: Summary of systematic errors.

sample into twenty pairs of disjoint sets. We measure the mass for each of the disjoint sets and take the RMS of the difference between them as an estimate of the error. Summing these, we get $0.5 \text{ GeV}/c^2$. We also estimate an uncertainty coming from possible imperfections in modeling the two largest sources of background: Drell-Yan and events with a “fake” lepton. This uncertainty is estimated to be $0.2 \text{ GeV}/c^2$.

Finally, the uncertainties in the parton distribution functions (PDFs) are estimated by using different PDF sets (CTEQ5L vs. MRST72), different values of Λ_{QCD} , and varying the eigenvectors of the CTEQ6M set, yielding a total uncertainty of $0.6 \text{ GeV}/c^2$.

VI. RESULT IN DATA

We apply the procedure described in Section III to the 344 candidate events observed in the data. After applying the corrections described in Section IV, we measure a top quark mass of

$$M_{top} = 171.2 \pm 2.7(\text{stat.}) \text{ GeV}/c^2$$

The final posterior probability density for the events in data can be seen in Figure 6.

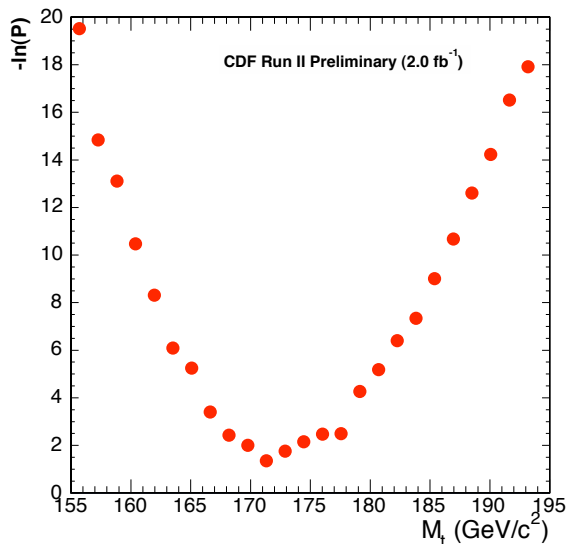


FIG. 6: Final posterior probability density as a function of the top pole mass for the 344 candidate events in data.

The measured statistical uncertainty is consistent with that measured for Monte Carlo experiments using $M_{top} = 172 \text{ GeV}/c^2$ signal events as shown in Figure 7.

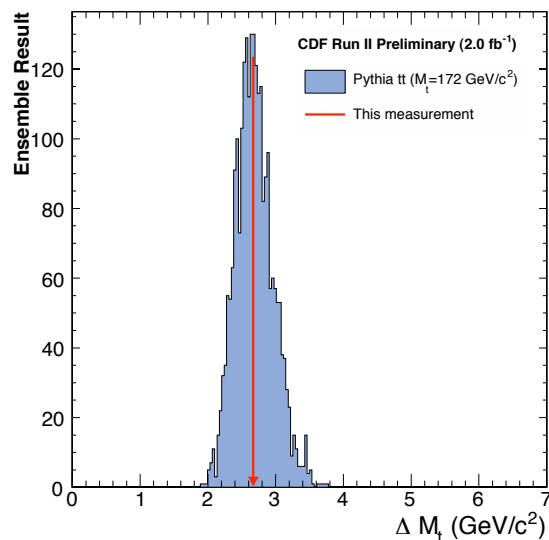


FIG. 7: Distribution of expected errors for $M_t = 172 \text{ GeV}/c^2$. The measured error is shown as the line; 53% of Monte Carlo experiments yielded a smaller error.

VII. CONCLUSION

We measure the top quark mass to be

$$M_{top} = 171.2 \pm 2.7(\text{stat.}) \pm 2.9(\text{syst.}) \text{ GeV}/c^2$$

in dilepton events in 2.0 fb^{-1} of CDF II data. We have used a normalized per-event differential cross-section for leading order top quark pair production and background to form a posterior probability. We utilize a neural network-based selection optimized for precision in measurement of the top quark mass.

Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comisión Interministerial de Ciencia y Tecnología, Spain; in part by the European Community's Human Potential Programme under contract HPRN-CT-2002-00292; and the Academy of Finland.

-
- [1] B. Abbott *et al.*, (DØ Collaboration), Phys. Rev. Lett. **80**, 2063 (1998).
 - [2] F. Abe *et al.*, (CDF Collaboration), Phys. Rev. Lett. **82**, 271 (1999).
 - [3] K. Kondo, J. Phys. Soc. Jpn. **57**, 4126 (1988).
 - [4] J. Estrada, FERMILAB-THESIS-2001-07 (2001).
 - [5] F. Canelli, FERMILAB-THESIS-2003-22 (2003).
 - [6] The CDF Collaboration, CDF/TOP/PUB/7718.
 - [7] A. Abulencia *et al.*, The CDF Collaboration, Phys. Rev. Lett. **96**, 152002 (2006).
 - [8] A. Abulencia *et al.*, The CDF Collaboration, Phys. Rev. D **74**, 032009 (2006).
 - [9] The CDF Collaboration, CDF/TOP/PUB/8951.
 - [10] D. Acosta *et al.*, The CDF Collaboration, Phys. Rev. Lett. **93**, 142001 (2004).

- [11] A. Abulencia *et. al.*, The CDF Collaboration, Phys. Rev. D **75**, 031105(R) (2007).
- [12] D. Acosta, et al., The CDF Collaboration, Phys. Rev. D **71**, 032001 (2005).
- [13] S. Whiteson and D. Whiteson. “Stochastic optimization for collision selection in high energy physics.” In *IAAI 2007: Proceedings of the Nineteenth Annual Innovative Applications of Artificial Intelligence Conference*, pages 1819-1825, July 2007.
- [14] K. O. Stanley and R. Miikkulainen, Evolutionary Computation **10(2)** 99-127, 2002.
- [15] G. Mahlon, S. Parke, Phys. Lett. B **411**, 173 (1997).
- [16] G. Mahlon, S. Parke, Phys. Rev. D **55**, 7249 (1996).